AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims

1. - 11. (Canceled)

12. (Currently Amended) A computer-implemented method for inferring genomic sequences unique to at least one set of organisms other than a set of organisms under investigation, the method comprising:

obtaining genomic data characteristic of a set of organisms under investigation;

formatting by a computer the genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a similarity search engine;

searching a selected genomic computer database using the query-length sequence and the similarity search engine for those sequences having homology above a threshold with at least one set of organisms other than the set under investigation,

the selected genomic computer database containing genomic data from a plurality of organisms;

parsing results of the search ~~for those sequences having homology above a threshold with at least one set of organisms other than the set under investigation and otherwise~~ for a sequence unique ~~within~~ to the selected genomic database~~;~~ and

outputting by the computer to a user an identity of those sequences ~~having homology above a threshold with at least one set of organisms other than the set under investigation and otherwise~~ unique ~~within~~ to the selected genomic database.

13. (Currently Amended) A computer-implemented system ~~program product~~ for inferring genomic sequences unique to a at least one set of organisms other than a set of organisms under investigation, the computer-implemented system ~~program product~~ comprising:

a computer-readable physical medium;

a genomic data interface module, stored on the medium and operable to couple to a source of genomic data to receive genomic data characteristic of a set of organisms under investigation;

a formatting module, stored on the medium and operable to format received genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a similarity search engine;

a search interface module, stored on the medium and operable to interface with the similarity search engine to submit the query-length sequence to a selected genomic database containing genomic data from a substantial plurality of organisms; and

a search results parsing module, stored on the medium and operable to parse results of the search for those sequences having homology above a threshold with at least one set of organisms other than the set under investigation and otherwise unique within the selected genomic database, and to output to a user an identity of those sequences having homology above a threshold with at least one set of organisms other than the set under investigation and otherwise unique within the selected genomic database.

14. (Currently Amended) A computer-implemented method for inferring genomic sequences unique to a first set of organisms, the method comprising:

obtaining genomic data characteristic of a second set of organisms;

formatting by a computer the second set genomic data into at least one query-length sequence, each query-length sequence being of a format compatible with a similarity search engine;

searching by the computer a selected genomic database using the query-length sequence and the similarity search engine, the selected genomic database containing genomic data from a plurality of organisms including the first set of organisms;

parsing results of the search for those sequences, other than sequences of the second set, having homology above a threshold with the second set and otherwise unique within the selected genomic database;; and

outputting by the computer to a user an identity of those sequences having homology above a threshold with the second set and otherwise unique within the selected genomic database as genomic sequences unique to the first set.

15. (New) The computer implemented method according to claim 12, further comprising removing the sequences from the selected genomic database that are not unique.